

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number
WO 02/099129 A2

(51) International Patent Classification⁷: C12Q 1/68

(74) Agent: CAMPBELL, Patrick, John, Henry; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5JJ (GB).

(21) International Application Number: PCT/GB02/02640

(22) International Filing Date: 7 June 2002 (07.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0113908.8 7 June 2001 (07.06.2001) GB

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(71) Applicant (*for all designated States except US*): UNIVERSITY COLLEGE LONDON [GB/GB]; Gower Street, London WC1E 9BT (GB).

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): GRIFFITHS, David, John [GB/GB]; Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF (GB). KELLAM, Paul [GB/GB]; Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF (GB). WEISS, Robert, Anthony [GB/GB]; Wohl Virion Centre, Windeyer Institute of Medical Sciences, University College London, 46 Cleveland Street, London W1T 4JF (GB).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: DESIGNING DEGENERATE PCR PRIMERS

(57) Abstract: A method of designing PCR primers for screening for new members of known virus families.



WO 02/099129 A2

DESIGNING DEGENERATE PCR PRIMERS

Field of the invention

The invention relates to a method of designing a panel of primers for detecting viruses in a high throughput polymerase chain reaction assay.

Background of the invention

5 All organisms appear to be capable of infection by viruses, including bacteria, animals and plants. Viruses require the use of the cellular translation and transcription machinery to replicate. In the process of replication they often have deleterious effects on the host cell and thus on the host organism. Viruses constitute an important class of pathogens causing many diseases, leading to loss of life in
10 humans and economic loss in the agricultural industries.

Summary of the invention

The polymerase chain reaction (PCR) allows the amplification of a specific region of a polynucleotide. The specificity of the reaction is due to the primers
15 which during the course of PCR bind to the region to be amplified in a sequence specific manner. The invention provides a method of designing primers which can be used in high throughput screening to detect viruses. The method may be used to detect unknown viruses which have not yet been sequenced.

In particular the invention provides a method of designing a panel of
20 degenerate primer pairs for screening for new members of multiple known virus families in a biological sample, wherein each primer pair in the panel binds a sequence that is conserved across members of a said virus family and selectively directs amplification of sequence of said family by PCR, which method comprises

- (a) providing a plurality of amino acid sequences from members of a first
25 virus family,
- (b) comparing the sequences to identify conserved regions,
- (c) designing a first primer pair using a computer based method, wherein each primer in the pair binds a nucleotide sequence that encodes a conserved region

-2-

identified in (b) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the nucleotide sequences that encode conserved regions in members of the first virus family, and

(d) repeating steps (a) to (c) for each virus family.

5 The invention also provides a method of designing a panel of degenerate primer pairs for screening for new members of multiple known virus families in a biological sample, wherein each primer pair in the panel binds a sequence that is conserved across members of a said virus family and selectively directs amplification of sequence of said family by PCR, which method comprises

10 (a) providing a plurality of nucleotide sequences from members of a first virus family,

(b) comparing the sequences to identify conserved regions,

(c) designing a first primer pair using a computer based method, wherein each primer in the pair binds a conserved region identified in (b) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the conserved regions in members of the first virus family, and

(d) repeating steps (a) to (c) for each virus family.

The invention additionally provides a panel of primers which has been designed by the method of the invention.

20

Detailed description of the invention

The invention provides a method of designing a panel of primer pairs which can be used in high throughput virus screening. The method comprises initial steps which deduce the sequences of the primers using computer based calculations, and optional later steps in which the primers are synthesised and tested empirically, for example to determine optimal PCR conditions and/or to select primer pairs with desired further properties.

25 The panel of primers provided by the method are designed to be capable of detecting unknown viruses based on nucleotide and/or amino acid sequences in the unknown virus which are similar (homologous) to nucleotide and/or amino acid sequences in a known virus. These conserved sequences typically have a role in

30

-3-

providing a necessary or advantageous activity or property to the virus. Conserved nucleotide sequences may be coding or non-coding sequences.

In one embodiment the conserved sequences code for or are from virus proteins which have the following activities: DNA or RNA polymerase (replicase),
5 topoisomerase (helicase/gyrase), endonuclease (integrase), nucleic acid binding protein, protease, transcription factors, envelope glycoproteins, structural protein (e.g. capsid or nucleocapsid protein).

The panel of primers is designed to detect viruses which are single stranded or double stranded DNA or single stranded or double stranded RNA viruses. The
10 viruses are generally capable of infecting prokaryotic or eukaryotic cells, such as bacterial, animal, plant, yeast or fungal cells. Preferably the viruses are mammalian (preferably primate) or avian viruses, such as human, pig, horse, sheep, goat, cow, chicken, turkey or duck viruses.

The viruses are typically from any combination of the following families:
15 Adenoviridae, Arenaviridae, Arteriviridae, Astroviridae, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Deltavirus, Filoviridae, Flaviviridae, Hepadnaviridae, Herpesviridae, Orthomyxoviridae, Papovaviridae, Paramyxoviridae, Parvoviridae, Picornaviridae, Polydnviridae, Poxviridae, Reoviridae, Retroviridae, Rhabdoviridae, Togaviridae or Bornavirus.

20 The primers of the panel are capable of detecting unknown viruses in a biological sample. Such a sample either originates from a single individual or is a pooled sample from individuals of the same species. Thus the panel of primers detects viruses which infect the same species (from which the sample originates).

Generally in the method at least 15, 30, 50, 100, 200 or more, typically up to
25 a maximum of 300 different primer pairs are designed. The primer pairs designed in the method bind sequence which is conserved across members of a virus family. The panel which is designed in the method may comprise primer pairs that bind sequence which is conserved across substantially all members of the family or across a subset of the members of the family, for example across all members of a subfamily or of a
30 genus. Generally, the primer pairs bind at least 70%, at least 80%, or at least 90% of the known viruses of the family, subfamily or genus. Preferably less than 10, such as

less than 5, primer pairs will be used for the detection of any given family, subfamily or genus in the panel.

The panel of primer pairs is generally capable of detecting viruses from at least 10, 15, 20, 30 or more families, typically up to a maximum of 35 families.

5 The panel of primer pairs may comprise sets of primer pairs which perform a nested PCR reaction. Generally such a set of primer pairs comprises a first and second primer pair. The first primer pair is able to amplify a template nucleotide sequence from a virus to form a PCR product. The second primer pair is able to amplify a nucleotide sequence using the PCR product generated by the first primer
10 pair as a template. The use of nested sets of primer pairs allows increased sensitivity.

In a preferred embodiment each primer pair is specific for a particular virus family, so that it does not detect viruses of other families.

In the method of the invention the plurality of amino acid sequences or nucleotide sequences are provided from different known viruses of the same family.

15 The amino acid sequences or nucleotide sequences will be for the same protein of the different viruses. Typically at least 5, 10, 20, 50, 100 or more sequences are provided. The maximum number of sequences provided will, for example, be 300 sequences.

Each of the sequences which is provided is typically at least 20, 50, 100, 200
20 or more amino acids or nucleotides in length. In general the maximum length of the nucleotide sequences is 1000 nucleotides and the maximum length of the amino acid sequences is 300 amino acids. The sequences may be obtained from a database of sequences, such as GenBank. The sequences may be obtained from a database comprising virus sequences which are organised into homologous protein families
25 (based on sequence similarity relationships).

In a preferred embodiment the sequences are obtained from the VIDA database (described in Alba et al (2001) Nucleic Acids Research 29, 133-136) or the Virus Division of GenBank. The sequences may be provided in the form of a database, preferably in computer-readable form. The sequences are preferably
30 provided in the form of a computer-readable database constructed using programs which identify homologous protein families, such as GeneTableMaker, MKDOM or

PSCBuilder.

The sequences which have been provided are compared to identify conserved regions. Typically such conserved regions will have a length of at least 12 nucleotides, such as at least 15, 21, 27, 36, 99 or more nucleotides (generally up to a maximum length of 200 nucleotides) or at least 4, 5, 7, 10, 25 or more amino acids (generally up to a maximum length of 50 amino acids).

Across the conserved region the virus sequences which are being provided will of course share identity or similarity. Typically the amino acids or nucleotides in at least 50% of the positions in the region will be the same in at least 50 %, 60%, 70%, or 80% of the viruses of the group (i.e. in the family, genus or subfamily).

The algorithm which identifies conserved regions generally uses a multiple sequence alignment method. The method may comprise (a) aligning all pairs of sequences separately to calculate a distance matrix giving the divergence of each pair of sequences, (b) calculating a guide tree from the distance matrix, and (c) aligning the sequences progressively according to the branching order in the guide tree.

A preferred algorithm for the aligning the conserved sequences is CLUSTALW as described in Thompson et al (1994) Nucleic Acids Research 22, 4673-80. Other algorithms that can be used for aligning sequences are MultAlin (Corpet (1988) Nucleic Acids Research 16, 10881-90) or Jalview (Clamp et al (1998) <http://barton.ebi.co.uk>). BLOCKS of conserved regions of amino acids may be extracted from the multiple alignments, typically using the program Blocks Multiple Alignment Processor. Alternatively the entire process of performing multiple alignments and extracting BLOCKS can be performed using BLOCKMAKER (Henikoff and Henikoff (1994) Genomics 19, 97-107).

The output from the alignment and BLOCK extraction set (i.e. the information describing the identified conserved regions) is then entered into the algorithm which designs the primers. Such output is typically in the form of partial sequences which correspond to the conserved regions (BLOCKS). These BLOCKS are input into a primer design algorithm. In one embodiment such an algorithm is CODEHOP.

In the primer design step the conserved regions which are chosen as targets

-6-

for primers preferably comprise few codons with degenerate counterparts, i.e. preferably the sequence has a low redundancy, such as a redundancy of less than 512 fold, 256 fold or 128 fold. Each primer binds in accordance with Watson-Crick base pairing and thus the binding is sequence specific. Each primer will thus be designed
5 to be wholly or partially complementary to the sequence to which it binds.

Each of the primers typically has a length of at least 8 nucleotides, such as at least 10, 12, 15, 20, 30, 40 or more nucleotides (up to a maximum of 50 nucleotides for example). In one embodiment the primer may comprises at least 2, 4 or 6, up to a maximum of 10 for example, inosine bases. Inosine is able to bind to any of the four
10 nucleotides and therefore use of inosine causes a reduction in effective redundancy.

Each primer pair will be designed so that the PCR product generally has a length of at least 20, such as at least 50, 100, 200, 500, 1000 or more nucleotides (and typically up to a maximum of 5×10^3 nucleotides long).

Each primer is preferably designed so that it anneals to a single site, i.e. the
15 primer will not bind to any other site in the genome of the relevant virus.

Each primer is preferably designed so that it does not exhibit secondary structure, i.e. the nucleotides in the primer will not bind substantially to any other nucleotide in the primer apart from those to which it is covalently linked. In addition preferably each primer is designed so that it does not bind other primers with the
20 same sequence.

In one embodiment the 3' region, and preferably the 3' terminal nucleotide of the primer binds to the target sequence with high affinity, thus preferably this region or nucleotide comprises a G or C.

Generally each primer is designed to have an annealing temperature of from
25 30 to 65°C, such as 50 to 60°C or 35 to 45°C. In addition each primer pair may be designed to ensure that the two primers do not bind to each other.

The primers are designed by a computer based algorithm. In one embodiment such an algorithm designs primers according to the following rules:

1) A set of blocks is input, where a block is an aligned array of amino acid
30 sequence segments without gaps that represents a highly conserved region of homologous proteins. A weight is provided for each sequence segment, which can be

-7-

increased to favour the contribution of selected sequences in designing the primer. A codon usage table is chosen for the target genome.

2) An amino acid position-specific scoring matrix (PSSM) is computed for each block using the odds ratio method.

5 3) A consensus amino acid residue is selected for each position of the block as the highest scoring amino acid in the matrix.

4) For each position of the block, the most common codon corresponding to the amino acid chosen in step 3 is selected utilizing the user-selected codon usage table. This selection is used for the default 5' consensus clamp in step 8.

10 5) A DNA PSSM is calculated from the amino acid matrix (step 2) and the codon usage table. The DNA matrix has three positions for each position of the amino acid matrix. The score for each amino acid is divided among its codons in proportion to their relative weights from the codon usage table, and the scores for each of the four different nucleotides are combined in each DNA matrix position.

15 Nucleotide positions are treated independently when the scores are combined. As an option, the highest scoring nucleotide residue from each position can replace the most common codons from step 4 that are used in the consensus clamp.

6) The degeneracy is determined at each position of the DNA matrix based on the number of bases found there. As an option, a weight threshold can be specified
20 such that bases that contribute less than a minimum weight are ignored in determining degeneracy.

7) Possible degenerate core regions are identified by scanning the DNA matrix in the 3' to 5' direction. A core region must start on an invariant 3' nucleotide position, have length of 11 or 12 positions ending on a codon boundary, and have a
25 maximum degeneracy of 128 (this is the default setting of CODEHOP). The degeneracy of a region is the product of the number of possible bases in each position.

8) Candidate degenerate core regions are extended by addition of a 5' consensus clamp from step 4 or 5. The length of the clamp is controlled by a melting
30 point temperature calculation (the CODEHOP default is 60°C) and is usually about 20 nucleotides.

-8-

9) Steps 7 and 8 are repeated on the reverse complement of the DNA matrix from step 5 for primers corresponding to the opposite DNA strand.

In one embodiment CODEHOP (Rose et al (1998) Nucleic Acids Research 26, 1628-1635) is used to design the primer pairs. This program uses the above
5 rules.

The primers designed by the algorithm may then be mapped back to the original sequence to choose primer pairs which provide the desired length of PCR product.

The above-described computer based method is repeated until the desired
10 number of primer pairs have been designed. Optionally the primer pairs can then be synthesised and tested. They are typically tested to determine the optimal conditions for using the primers in a PCR reaction.

The PCR reaction is carried out in a PCR mixture that generally comprises the following: the template polynucleotide (which will be amplified in the event of
15 virus detection), one or more primer pairs designed as described above, a polymerase enzyme (typically a DNA polymerase, such as Taq polymerase), deoxynucleotide triphosphates (dATP, dTTP, dCTP and dGTP) and a suitable buffer.

The PCR reaction generally comprises cycles of the following steps: a denaturation step, a primer annealing step and a polynucleotide synthesis step.
20 Typically the PCR reaction comprises at least 25 cycles, such as 30, 35, 40 or more cycles, up to a maximum of 60 cycles for example. Generally in the denaturation step the PCR mixture is heated to a temperature at which the polynucleotides in the PCR mixture (in particular the polynucleotide region to be amplified) denature to single stranded form. The denaturing temperature is generally from 85 to 98°C.

25 In the primer annealing step the primers bind to template nucleotide sequence in a sequence specific manner. This step is generally carried out at a temperature of from 30 to 65°C. In the polynucleotide synthesis step the polymerase replicates/synthesises nucleotide sequence based on template sequence by addition of nucleotides to the 3' end of the bound primers. This step is generally carried out at
30 about 72°C.

In one embodiment the primers are tested for their ability to amplify one or

more of the plurality of nucleotide sequences from known viruses which were used to design the primers, or in the case of amino acid sequences from known viruses being used to design the primers the primers may be tested for their ability to amplify the nucleotide sequence from the virus which encodes the amino acid sequence.

- 5 The primers may be tested in a range of buffer conditions to determine optimal buffer conditions for PCR using the primers. The buffer conditions which may be tested include pH (typically between 7 and 10), magnesium concentration (typically from 0.5 mM to 5 mM), potassium chloride (typically from 0 to 100 mM), ammonium chloride (typically 0 to 100 mM), glycerol (typically 0 to 20%),
10 dimethylsulphoxide (typically 0 to 20%), ethanol (typically 0 to 20%), sorbitol (typically 0 to 20%) or betaine (typically 1M betaine).

The primers may be tested at a range of different temperatures to determine the optimal temperatures in the PCR reaction. Preferably the primers are tested in PCR reaction in which a range of primer annealing temperatures are tested.

- 15 Typically the range of temperatures is from 30 to 65°C.

The panel of primer pairs or a group of primers within the panel may be designed to be used together on the same plate (i.e. using the same thermal cycles). Thus such primer pairs will be designed to work at the same annealing temperature.

- In one embodiment a group of primer pairs within the panel are designed to
20 have similar optimal conditions for use in PCR so that they can be used optimally in the same well or reaction vessel, i.e. that they can be used in multiplex PCR. Such a group typically comprises at least 2, 3, 4, 5, 6 or more primer pairs (up to a maximum of 8 primer pairs for example).

- To provide such primer pairs the computer based method steps may be used
25 to design primer pairs which are calculated to have similar annealing temperatures and/or the primers are tested to select primer pairs which can be used optimally together. Such testing typically determines whether the primers work optimally with the same buffers and/or whether the primers have similar annealing temperatures.

- In one embodiment at least one or both primers of each primer pair in the
30 group carries a label. Typically at least one of the primers in each primer pair will carry a different label from that used for the other primer pairs. The PCR product

generated by labelled primers carries the labels present on the primers. Thus after the group of primers have been used for PCR in the same well detection of the labels in the PCR products can be used to deduce which PCR product was formed from each primer pair. In one embodiment all forward primers of the group are labelled with one colour and the reverse primers are labelled with a different colour.

In a preferred embodiment the primers are labelled with a fluorescent label, such as fluorescein based labels (e.g. fluorescein isothiocyanate). Different primer pairs may be labelled with fluorescent labels of different colours. The fluorescent labels which are used may be capable of detection by a Beckman CEQ2000™ or Applied Biosystems A3700™ fluorescent DNA analyser. The fluorescent labels may be obtained from Beckman Coulter or Applied Biosystems.

Another way of being able to determine which PCR products are generated by which primer pair is for each primer pair in the group to generate a PCR product of different size to the PCR products generated by the other primer pairs of the group. Typically each PCR product which is generated by the group of primers differs in size from all the other PCR products by at least 20, such as at least 50, 100, 200, 500, 1000 or more nucleotides. Each PCR product may for example differ in size from all other PCR products by up to a maximum of 3000 nucleotides.

The following Example illustrates the invention:

20 Example

The Example below refers to Figure 1 which shows how primers were designed using a database known as 'VIDA', and computer programs known as 'CLUSTALW', 'BLOCKMAKER' (or 'BLOCKS') and 'CODEHOP'.

25 Designing a panel of primers

A panel of primers was designed for detecting unknown viruses from the family Herpesviridae according to the strategy shown in Figure 1. The amino acid sequences of herpes virus DNA packaging protein UL15 were obtained from the VIDA database (Alba et al, see above). These sequences are shown in Table 1.

30 The sequences obtained from the VIDA database were then imported into CLUSTALW. This compares the protein sequences to identify conserved regions

-11-

and then aligns the sequences according to the conserved regions. The alignment produced by CLUSTALW is shown in Table 2.

The BLOCKMAKER program was then used to extract blocks of conserved aligned sequences which do not contain gaps from CLUSTALW and enter them into
5 CODEHOP. The primer sequences were then designed by CODEHOP using the conserved sequences. The output from the CODEHOP program is shown in Table 3. The 'Complement of Block' sequences shown in Table 3 shows the sequence of the other strand allowing primers to be designed for amplification in the opposite direction.

10

Table 1. All protein sequences of DNA packaging protein UL15 extracted from VIDA.
Here written as a list and unaligned.

```

>gi_10180719
5  MFGGLLGEETKRHFERLMKTNDRLGASHRNEIRDGMVDAPFLNFAIPVPRRHQTVMPAIGILHNCC
   DSLGIYSAITTRMLYSSIA CSEFDELRRDSVPRCYPRITNAQAFSPMMRVANSIIFQEYDEMECAHR
   NAYYSTMNSFISMRTSDAFKQLTVFISRFSKLLIASFRDVKLLDDHTVKKRARIDAPSYDKLHGTLELFQ
   KMILMHATYFVTSVLLGDHAERAERLLRVAFDTPHFSDIVTRHFRQRATVFLVPRRHGKTWFLVPLIALA
   MSSFEGIRIGYTSHIRKAIEPVFEDIGDRLRRWFGAHRVDHVKGETITFSFPSGLKSTVTFASSHTNSI
10  RGQDFNLLFVDEANFIRPDVQTIIGFLNQATCKIIFVSSTNSGKASTSFLYGLKGSADDLLNVVTYICD
   EHKHVTDYTNATSCSCYVLNKPVFITMDGAMRRTAEMFLPDSFMQEIIGGGVVDRTICQGDRSIFTASA
   IDRFLIYRPSTVNNQDPFSQDLYVYVDPAPTANTKASGTGVAVIGKYGTDYIVFGLHEHYFLRALTGESSD
   SIGYCVAQCLIQICAIHRKRFGVIKIAIEGNSNQDSAVAIATRIAIEMISYMKA AVAPT PHNVSFYHSKS
   NGTDVEYPYFLLQRQKTTAFDFFIAQFN SGRVLASQDLVSTTVSLTDPVEYLTQKLTNISEVVTGPTCT
15  RTFSGKKGGNDDTVVALTMVYISAHIPDMAFAPIRV

>gi_7673189
   MFGGLLGEETKRHFERLMKTNDRLGASHRNEIRDGMVDAPFLNFAIPVPRRHQTVMPAIGILHNCC
   DSLGIYSAITTRMLYSSIA CSEFDELRRDSVPRCYPRITNAQAFSPMMRVANSIIFQEYDEMECAHR
   NAYYSTMNSFISMRTSDAFKQLTVFISRFSKLLIASFRDVKLLDDHTVKKRARIDAPSYDKLHGTLELFQ
20  KMIFDACHLFCNFCFTWRSRRASERLLRVAFDTPHFSDIVTRHFRQRATVFLVPRRHGKTWFLVPLIALA
   MSSFEGIRIGYISHIRKAIEPVFEDIGDRLRRWFGAHRVDHVKGETITFSFPSGLKSTVTFASSHTNSI
   RGQDFNLLFVDEANFIRPDVQTIIGFLNQATCKIIFVSSTNSGKASTSFLYGLKGSADDLLNVVTYICD
   EHKHVTDYTNATSCSCYVLNKPVFITMDGAMRRTAEMFLPDSFMQEIIGGGVVDRTICQGDRSIFTASA
   IDRFLIYRPSTVNNQDPFSQDLYVYVDPAPTANTKASGTGVAVIGKYGTDYIVFGLHEHYFLRALTGESSG
25  SIGYCVAQCLIQICAIHRKRFGVIKIAIEGNSNQDSAVAIATRIAIEMISYMKA AVAPT PHNVSFYHSKS
   NGTDVEYPYFLLQRQKTTAFDFFIAQFN SGRVLASQDLVSTTVSLTDPVEYLTQKLTNISEVVTGPTCT
   RTFSGKKGGNDDTVVALTMVYISAHIPDMAFAPIRV

>gi_5689285
   MFGGALGESAKKHFERLLRDRNERLGASRKN ECLARGGSLVDAPFLNFAISVPRRHQTVMPAVGTLHDCC
30  DGTGIYSAIATRLLYAGIVSSEFGEVRRRESLSNGHISKRNREALLAPTLTRVANSITFHEYDDAQCAHR
   NAYYSTMTNFTGSMRTSDAFQQLASFIDRF SKLLAASF KDVNILDRNNAPKRARITAPSYDKPHGTLELFQ
   KMILMHATYFLTSLLEDHAERAERLLRVIFDIPDFSDAATRHFRQRATVFLVPRRHGKTWFLVPLIALA
   MSSFEGIRIGYTSHIRKAIEPVFEEIGDRLRRWFGTQCVDHVKGETITFSFPSGSRSTVTFASSHTNSI
   RGQDFNLLFVDEANFIRPDVQTIIGFLNQANCKIIFVSSTNSGKASTSFLYGLKGSADDLLNVVTYICD
35  EHKHVTNYTNATSCSCYVLNKPVFITMDGAMRRTAEMFLPDSFMKEIIGGITMDRNTCQGD RGVTASA
   VERLLLYRPSTVRNQDILSRDLYVYVDPAPTANTRASGTGIAVIGRYGADYIIFGLEHFFLRALTGESAD
   AIGECAAQCIAQICAIHCERFGTIRVAVEGNSNQDSAVAIATRISIDLASYVQSGVAPAPHDVCFYH SKP
   AGSNVEYPFFLLQRQKTAADFDFIARFN SGRVLASQDLVSTTISLSTDPVEYLTQKLTNLSEVVTGATGT
   RTFSGKKGGYDDTVVALVMVYISAHASDATFAPIRGVEATCRGPTEA
40  >gi_1869837
   MFGQQLASDVQQYLERLEKQRQQKVGVD EASAGLT LGGDALRVFFLDFATATPKRHQTVVPGVGT LHDCC
   EHSPLFSAVARRLLFNLSLVAQLRGRDFG GDHTAKLEFLAPELVRVARLRFRECAPEDAVPQRNAYYSV
   LNTFQALHRSEAFRQLVHFVRDFAQLL KTSFRASSLAETTGPPEKKRAKVDVATHGQTYGTLELFQ KMILM
   HATYFLAAVLLGDHAEQVNTFLRLVFEI PLFSDTAVRHFRQRATVFLVPRRHGKTWFLVPLIALSLASFR
45  GIKIGYTAHIRKATEPVFDEIDACL RGFSGSRVDHVKGETISFSFPDGSRSSTIVFASSHTNGIRGQDF

```

-13-

NLLFVDEANFIRPDVQTIMGFLNQANCKIIFVSSNTNGKASTSFLYNLRGADELINNVVTYICDDHMPR
 VVTHTNATACSCYILNKPVFITMDGAVRRATDLFLPDSFMQEIIGGQARETGDDRPVLTSAKERFLLYR
 PSTTTNSGLMAPELYVYVDPFAFTANTRASGTGIAVVGGRYRDDFIIFALEHFFLRALTGSAPADIARCVVH
 SLAQVLALHPGAFRSVRVAVEGNSSQDSAVAIAITHVHTEMHRILASAGANGPGPELLFYHCEPPGAVLY
 5 PFFLLNKQKTPAFEFYIKKFNSGGVMASQELVSVTVRLQTDPEYLLSEQLNNLIETVSPNTDVRMYSGKR
 NGAADDLMVAVIMAIYLAAPTGIIPAFFPITRTS
 >gi_59501
 MFGQQLASDVQQYLERLEKQRLKVGADSEASAGLTMGGDALRVPFLDFATATPKRHQTVVPGVGTLLHDC
 EHSPLFSAVARRLLFNSLVPAQLKGRDFGDDHTAKLEFLAPELVRAVARLRFKECAPADVVPQRNAYYSV
 10 LNTFQALHRSEAFRLVHFVRDFAQLLKTSFRASSLTETTGPCKKRAKVDVATHGRTYGTLELFQKMILM
 HATYFLAAVLLGDHAEQVNTFLRLVFEIPLFSDAAVRHFRQRATVFLVPRRHGKTWFLVPLIALSLASFR
 GIKIGYTAHIRKATEPVFEEIDACLRGWFSGARVDHVKGETISFSFPDGSRTIVFASSHNTNGIRGQDF
 NLLFVDEANFIRPDVQTIMGFLNQANCKIIFVSSNTNGKASTSFLYNLRGADELINNVVTYICDDHMPR
 VVTHTNATACSCYILNKPVFITMDGAVRRATDLFLADSFQEIIGGQARETGDDRPVLTSAKERFLLYR
 15 PSTTTNSGLMAPDLYVYVDPFAFTANTRASGTGVAVVGGRYRDDYIIFALEHFFLRALTGSAPADIARCVVH
 SLTQVLALHPGAFRGVRVAVEGNSSQDSAVAIAITHVHTEMHRLASEGADAGSGPELLFYHCEPPGSAVL
 YPFFLLNKQKTPAFEFYIKKFNSGGVMASQEIIVSATVRLQTDPEYLLLEQLNNLTETVSPNTDVRTYSGK
 RNASDDLMVAVIMAIYLAQAQAGPPHTFAPITRVS
 >gi_2605992
 20 MFGKALSRETIQYFETLRKEVQSRSGAKNRAEAQGTGGEDDVKTAFLNFAIPTPQRHQTVPVPGVGTLLHDC
 CETAQIFASVARLLFRSLSKWRGGESKERLDPSSVEAYVDPKVKQALKTISFVEYNDAEARSERNAYYS
 IMNTFDSLRSDDAFHQVANFVARFSRLVDTSFNGADLDGDDGQTSKRIKVDVPTYGKQRTLELFQKMIL
 MHATYFIAAVILGDHADRIAGFLKMVFNTPEFSDATIRHFRQRATVFLVPRRHGKTWFLVPLIALALATF
 KGIKIGYTAHIRKATEPVFDEIGARLRQWFGNSPVDHVKGENISFSFPDGSKSTIVFASSHNTNGIRGQD
 25 FNLLFVDEANFIRPEAVQTIIGFLNQTNCKIIFVSSNTNGKASTSFLYNLRGADELINNVVTYICDEHME
 RVKAHTNATSCSCYILNKPVFITMDGAMRNTAELFLPDSFMQEIIGGNGISGAHRDEPVFTKTAQDRFLL
 YRPSTVANQDIMSNLYVYVDPFAFTNAMASGTGVAVVGGRYRSNWIIFGLEHFFLSALTGSSAELIARCV
 AQCLAKVFAIHSRPFDSVRVAVEGNSSQDAVAIATNIQLELNTLRQADVVMHMPGTVLFYHCTPPGSSVA
 YPFFLLQKQKTGAFDHFIFAFNSGLVLASQELISNTVRLQTDPEYLLLTQMKNLTEVITGTSETRVFTGK
 30 RNASDDMLVALVMAYMASLPPTTNAFSSLSLQ
 >gi_330792
 MFGRLVGRETVQYFEALRREVQARRGAKNRAEAQNGGEDDAKTAFLNFAIPTPQRHQTVPVPGVGTLLHDC
 CETAQIFASVARLLFRSLSKWQSGEARERLDPASVEAYVDPKVRQALKTISFVEYSDDEARSERNAYYS
 IMNTFDALRSDDAFHQVASFVARFSRLVDTSFNGADLDGDDGQASKRARVDVPTYGKQRTLELFQKMIL
 35 MHATYFIAAVILGDHADRIAGFLKMVFNTPEFSDATIRHFRQRATVFLVPRRHGKTWFLVPLIALALATF
 KGIKIGYTAHIRKATEPVFDEIGARLRQWFGNSPVDHVKGENISFSFPDGSKSTIVFASSHNTNGIRGQD
 FNLLFVDEANFIRPEAVQTIIGFLNQTNCKIIFVSSNTNGKASTSFLYNLRGAADDLLNVVTYICDEHME
 RVKAHTNATACSCYILNKPVFITMDGAMRNTAELFLPDSFMQEIIGGNGVSGAHRDEPVFTKTAQDRFLL
 YRPSTVANQDIMSSDLYVYVDPFAFTNAMASGTGVAVVGGRYRSNWWVFGMEHFFLSALTGSSAELIARCV
 40 AQCLAQVFAIHKRPFDSVRVAVEGNSSQDAVAIATNIQLELNTLRADVVPMPGAVLFYHCTPHGSSVA
 YPFFLLQKQKTGAFDHFIFAFNSGLVLASQELVSNVTVRLQTDPEYLLLTQMKNLTEVVTGTSETRVFTGK
 RNASDDMLVALVMAYLSSLPPTSDAFSSSLPAQ
 >gi_971317
 MFGGAVGEQSARYFORLLRERQRRAAERGAPDGGGGARGEDDARVPFLDFAVAAPKRHQTVPVPGVGTLLH
 45 GYCELAPLFAATASRLLLTSMARAEAGLNTGTGEAHVSRELAVLSALRFAAHPFAEAAAHNCNAYHVSMA
 ALESMRASGAFAQVAFAVARFSRLVGTFSHLLGGGDDADPPRAKRARVEPPSGQTRGALELFQKMILMPA

-14-

TYFVAATLLGEHAERIGAFRLVAFNTPDFSDAAVAHFRQRATVFLVPRRHGKTWFLVPLIALALATFKGI
KIGYTAHIRKATEPVFEEIIVARLRQWFGGERVDHVKGAVISFSFPD GARSTIVFASSHNTNGIRGQDFNL
LFVDEANFIRPEAVQTIVGFLNQASCKIIFVSSSTNTGKASTSFLYNLKGASDGLLNVTYICNEHTPRVA
AHGGATACSCYVLNKPVFITMDAAARNTAETFLPNSFMQEIIGGGEVARRAEPAAVFTRAAGEQFLLYRP
5 STAAARGPWPERLYMYIDPAFTSNARASGSGIAVVGHRGSLVLGLEHFFLPALTGSSAAEIARCAVRC
FAQVMAVHRRRLDGLFVAVEGNSSQDSAVAIALGVRRRLDLSAASGAVPMPAETRFYHCRPPGSAVAYPF
FLLQKQKTAADFHFIRLFNSGRVVASQDLASLTVRLQTDPEYVLFQQLNLTESTAGPGGARAFSGKRRG
AADDLMVALVMAVFGSLPPTDGAFCLAPRPPAD
>gi_5869808
10 MSLIMFGRTLGEESVRYFERLKRRLDERFGTLESPTPCSTRQGSIGNATQIPFLNFAIDVTRRHQAVIPG
IGTLHNCEYIPLFSATARRAMFGAFLSSTGYNCTPNVVLKPWRYSVNANVSPELKKAVSSVQFYEYSPE
EAAPHRNAYSGVMNTFRAFSLSDSFCQLSTFTQRFSLVETSFSFIEECGSHGKRAKVDVPIYGRYKGT
ELFQKMILMHTTHFISSVLLGDHADRVDCLRTVFNTSPVSDSVLEHFKQKSTVFLVPRRHGKTWFLVPL
IALVMATFRGKIVGYTAHIRKATEPVFEGIKSRLEQWFGANYVDHVKGESITFSFTDGSYSTAVFASHN
15 TNGIRGQDFNLLFVDEANFIRPDAVQTIVGFLNQTNCKIIFVSSSTNTGKASTSFLYNLRGSSDQLLNVT
YVCDHMPRVLASDVTACSCYVLNKPVFITMDGAMRRTADLFMADSFVQEI VGGKQKNSGGVGFDRPLF
TKTARERFILYRPSTVANCAILSSVLYVYVDPFTSNTRASGTGVAIVGRYKSDWII FGLEHFFLRALTG
TSSEIGRCVTQCLGHILALHPNTFTNVHVSIEGNSSQDSAVAISLAIAQQFAVLEKGNVLSAPVLLFY
HSIPPGCSVAYPFFLLQKQKTPAVDYFVKRFNSGNI IASQELVSLTVKLGVDPEYVLCQQLDNLTEVIK
20 GMGNLDTKTYTGKGTGTMSDDLMLVALIMSVYIGSSCIPDSVFMPIK
>gi_5708110
MLGKESVEIVKRYDALRKRRTMERGPDDVDGQEMSDSNFITTASICDRNDSARDTMNSPASRFQFAIDVP
QRHQACIAPIGSFHNCCAISRAFSYMASEIIYENLASYSTKYTDTDAALNDLQVSPKRQLFTGAAEDSIL
PALRQKLANLNFARFAPSDSLIHDKAFDGINNGYRGFVKSDEFSQLNNFIYRFHTLLKKSFGQASNDYK
25 RAKLEKTTSEQRDGTLELFQKMILMHATYFASSICLGEGSTERSNRYLSTVFNTSLFSENI IQHFRQRTT
VFLVPRRHGKTWFLVPLISLLVSSFEGIRIGYTAHLRKATEPVFIEIFTRLYKNFGAKQVEQVKGETITF
TFRNGKSAIVFASSQNTNGLRGQDFNLFVDEANFIKPAALHTVMGFLNQTNCKLFFVSSSTNTCHSNTS
LLYNLKGKTNLSLNVVTYICDEHMPETQKRTDVTTCSCYVLHKPVFVSMDSVRNTADLFVKDSFMHEIA
GGRAGKYDSDRTLVPVRALDQFLIYRPSTSSKPNISGLGKILTVYVDPFTTNRSASGTGIALVTALRDS
30 MVLGAEHFYLDALTGEAALEIAQCIVYLCIAYCCLIHAGAFREIRIAVEGNSSQDSAAAIAGNLTLLDS
LRRRLGFSLTFAHSRQPGTAMAHFFYLLNKQKSRAFDLFVSLFNSGRFMASQELVSNTLVLSKDPCEYLV
DQIRNITVTHGQGPDSFRFTSGKQGRVDDMLVAAMSTYLALEGSPTAGYHPAPIGRQRPA
>gi_1813970
MLRGDSAAKIQERYAELQKRKSHPTSCISTAFNTVATLCRKRYQMMHPGLAHS CNEAFLPLMAFCGRH
35 RDYNSPEESQRELLFHERLKSALDKLTFRPCSEEQRASQKLDALTELYRDPQFQQINNFMDFKKWLDG
GFSTAVEGDAKAI RLEPFQKNLLIHVIFFI AVTKIPVLANRVLQYLIHAFQIDFLSQTSIDIFKQKATVF
LVPRRHGKTWFIPIISFLLKHMIGISIGYVAHQHVSQFVLKEVEFRCHRTFARDYVVENKDNVISIDH
RGAKSTALFASCYNTNSIRGQNFHLLLVDEAHFIKKEAFNTILGFLAQNTTKIIFISSTNTSDSTCFLT
RLNNAPFMDLNVSVYCEEHLHSFTEKGDATACPCYRLHKPTFISLNSQVRKTANMFMGAFMDEIIGGT
40 NKISQNTVLITDQSEEFDILRYSTLNTNAYDYFGKTLVYVLDPAFTTNRKASGTGVAAVGAYRHQFLIY
GLEHFFLRDLSESSEVAIAECAAHMII SVLSLHPYLDELRIAVEGNTNQAAAVRIACLIRQSVQSSTLIR
VLFYHTPDQNHIEQPFYLMGRDKALAVEQFISRFNSGYIKASQELVSYTIKLSHDPIEYLLLEQIQNLHRV
TLAEGTTARYSAKRQNRISDDLIIAVIMATYLCDDIHAIRFRVS
>gi_2746296
45 MLRSCDIDAIQKAYQSI IWKHEQDVKISSTFPNSAIFCQKRFIILTPELGFTTHAYCRHVKPLYLFCDRQR
HVKSKIAICDPLNCALSCLKFTAIEKNTQVQYQKHLELQTSFYRNPMLQIEKFIQDFQRWICGFENT

-15-

NKKERIKLEPFQKSILIHIIFFISVTKLPTLANHVLDYLKYKFDIEFINESSVNILKQKASVFLVPRRHG
KTWFMIPVICFLLKNLEGISIGYVAHQKHVSFVMKDVEFKRRFFPQKNITCQDNVITIEHETIKSTAL
FASCYNTHSIRGQSFNLLIVDESHFIKKDAFSTILGFLPQSSTKIIFISSTNSGNHSTSFLTCLSNSPFE
MLTVSVSYVCEDHVHILNDRGNATTACCYRLHKPKFISINADVKKTADLFLEGAFKHEIMGGSCLCNVNDT
5 LITEQGLIEFDLFRYSTISKQIIPFLGKELYIIDPAYTINRRASGTGVAAIGTYGDQYIIYGMHEYFLE
SLLSNSDASIAECASHMILAVLELHPFFTELKIIIEGNSNQSSAVKIIACILKQTISVIRYKHITFFHTLD
QSQIAQPFYLLGREKRLAVEYFISNFGYIKASQELISFTIKITYDPVEYALEQIRNIHQININEHVTY
NAKKQTCSDDLLISIIIMAIYMCHEGKQTSFKEI
>gi_325496
10 MLRTCDITHIKNNEYAIIWKGERDCSTISTKYPNSAIFYKKRFIMLTPELGFAHSYNQVVKPLYTFCEKQ
RHLKNRKPLTILPSLSHLQEMKFLPASDKSFESQYTEFLESFKILYREPLFLQIDGFIKDFRKWIKGEF
NDFGDTRKIQLEPFQKNILIHVIFIAVTKLALANRVINYLTHVFDIEFVNESTLNTLKQKTNVFLVPR
RHGKTWFIPIISFLLKNIEGISIGYVAHQKHVSFVMKEVEFKRRMFPEKTITCLDNVITIDHQNKS
TALFASCYNTHSIRGQSFNLLIVDESHFIKKDAFSTILGFLPQASTKILFISSTNSGNHSTSFLMKLNN
15 SPFEMLSVVSVCEDHAHMLNERGNATACSCYRLHKPKFISINAEVKKTANLFLEGAFIHEIMGGATCNV
INDVLITEQGQTEFEFFRYSTINKNLIIPFLGKDLYVYLDPAYTGNRRASGTGIAAIGTYLDQYIIVGMEH
YFLESMTSSDTAIAECAAHMILSILDLPFFTEVKIIIEGNSNQASAVKIIACI IKENITANKSIQVTF
HTPDQONQIAQPFYLLGKEKKLAVEFFISNFGSNIKASQELISFTIKITYDPVEYALEQIRNIHQISVNN
YITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
20 >gi_854039
MKLNNSPFEMLSVVSVCEDHAHMLNERGNATACSCYRLHKPKFISINAEVKKTANLFLEGAFIHEIMGG
ATCNVINDVLITEQGQTEFEFFRYSTINKNLIIPFLGKDLYVYLDPAYTGNRRASGTGIAAIGTYLDQYIV
YGMHEYFLESMTSSDTAIAECAAHMILSILDLPFFTEVKIIIEGNSNQASAVKIIACI IKENITANKSI
QVTFHTPDQONQIAQPFYLLGKEKKLAVEFFISNFGSNIKASQELISFTIKITYDPVEYALEQIRNIHQ
25 ISVNNYITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
>gi_5733564
MLRTCDITHIKNNEYAIIWKGERNCSTISTKYPNSAIFYKKRFIMLTPELGFAHSYNQVVKPLYTFCEKQ
RHLKNRKPLTILPSLTRKLQEMKFLPASDKSFESQYTEFLESFKILYREPLFLQIDGFIKDFRKWIKGEF
NDFGDTRKIQLEPFQKNILIHVIFIAVTKLALANRVINYLTHVFDIEFVNESTLNTLKQKTNVFLVPR
30 RHGKTWFIPIISFLLKNIEGISIGYVAHQKHVSFVMKEVEFKRRMFPEKTITCLDNVITIDHQNKS
TALFASCYNTHSIRGQSFNLLIVDESHFIKKDAFSTILGFLPQASTKILFISSTNSGNHSTSFLMKLNN
PFEMLSVVSVCEDHAHMLNERGNATACSCYRLHKPKFISINAEVKKTANLFLEGAFIHEIMGGATCNVI
NDVLITEQGQTEFEFFRYSTINKNLIIPFLGKDLYVYLDPAYTGNRRASGTGIAAIGTYLDQYIIVGMEH
FLESMTSSDTAIAECAAHMILSILDLPFFTEVKIIIEGNSNQASAVKIIACI IKENITANKSIQVTFH
35 TPDQONQIAQPFYLLGKEKKLAVEFFISNFGSNIKASQELISFTIKITYDPVEYALEQIRNIHQISVNNY
ITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
>gi_4996048
MKLNNSPFEMLSVVSVCEDHAHMLNERGNATACSCYRLHKPKFISINAEVKKTANLFLEGAFIHEIMGG
ATCNVINDVLITEQGQTEFEFFRYSTINKNLIIPFLGKDLYVYLDPAYTGNRRASGTGIAAIGTYLDQYIV
40 YGMHEYFLESMTSSDTAIAECAAHMILSILDLPFFTEVKIIIEGNSNQASAVKIIACI IKENITANKSI
QVTFHTPDQONQIAQPFYLLGKEKKLAVEFFISNFGSNIKASQELISFTIKITYDPVEYALEQIRNIHQ
ISVNNYITYSAKKQACSDDLIIAIIIMAIYVCSGNSSASFREI
>gi_1136808
MLLSRHRERLANLEETAKDAGERWELSAPTFTRHCPKTARMAHPFIGVVHRINSYSSVLETYCTRHHPA
45 TPTSANPDVGTTPRPSEDNPAKPRLLESSTYLQMRVCREDAHVSTADQLVEYQAGRKTHDSLHACS VYR
ELQAFVLNLSFNLGCVPGVHWLEPFQQQLVMHTFFFLVSIKAPQKTHQLFGLFKQYFGLFETPNSVLQ

-16-

TFKQKASVFLIPRRHGKTWIVVAIIISMLLASVENINIGYVAHQKHVANSVFAEIIKTLCRWFPPKNLNLIK
 KENGTTIIYTRPGGRSSSLMCATCFNKNSIRGQTFNLLYVDEANFIKKDALPAILGFMLQKDAKLIFISSV
 NSSDRSTSFLNLNRNAQEKMLNVVSYVCADHREDFHLQDALVSCPCYRLHIPTYITIDESIKTTTNLFME
 GAFDTELMGEGAASSNATLYRVVGDAALTQFDMCRVDTTAQEVQKCLGKQLFVYIDPAYTNNTASGTGV
 5 GAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTTIERNVNAAVEGNSSQDSGVA
 IATVLNEICPLPIHFLHYTDKSSALQWPIYMLGGEKSSAFETFIYALNSGTLSASQTVVSNNTIKISFDPV
 TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDDLVAVVMMAHFMAATDDRMYKPISPQ
 >gi_1718281
 MLQKDAKLIFISSVNSSDRSTSFLNLNRNAQEKMLNVVSYVCADHREDFHLQDALVSCPCYRLHIPTYIT
 10 IDESIKTTTNLFMEGAFDTELMGEGAASSNATLYRVVGDAALTQFDMCRVDTTAQVQKCLGKQLFVYID
 PAYTNNTASGTGVGAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTTIERNV
 AAVEGNSSQDSGVAIATVLNEICPLPIHFLHYTDKSSALQWPIYMLGGEKSSAFETFIYALNSGTLSASQ
 TVVSNNTIKISFDPV TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDDLVAVVMMAHFMAATDDRMYKPISP
 Q
 15 >gi_2246515
 MLQKDAKLIFISSVNSSDRSTSFLNLNRNAQEKMLNVVSYVCADHREDFHLQDALVSCPCYRLHIPTYIT
 IDESIKTTTNLFMEGAFDTELMGEGAASSNATLYRVVGDAALTQFDMCRVDTTAQVQKCLGKQLFVYID
 PAYTNNTASGTGVGAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTTIERNV
 AAVEGNSSQDSGVAIATVLNEICPLPIHFLHYTDKSSALQWPIYMLGGEKSSAFETFIYALNSGTLSASQ
 20 TVVSNNTIKISFDPV TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDDLVAVVMMAHFMAATDDRMYKPISP
 Q
 >gi_2246552
 MLLSRHRERLAANLQETAKDAGERWELSAPTFTRHCPKTARMAHPFIGVVHRINSYSSVLETYCTRHPA
 TPTSANPDVGTTPRSEDNVPKPRLLLESSTYLQMCVREDAHVSTADQLVEYQAARKTHDSLHACSVYR
 25 ELQAFVLNLSFNGCYVPGVHWLEPFQQQLVMHTFFFLVSIKAPQKTHQLFGLFKQYFGLFETPNSVLQ
 TFKQKASVFLIPRRHGKTWIVVAIIISMLLASVENINIGYVAHQKHVANSVFAEIIKTLCRWFPPKNLNLIK
 KENGTTIIYTRPGGRSSSLMCATCFNKNSIRGQTFNLLYVDEANFIKKDALPAILGFMLQKDAKLIFISSV
 NSSDRSTSFLNLNRNAQEKMLNVVSYVCADHREDFHLQDALVSCPCYRLHIPTYITIDESIKTTTNLFME
 GAFDTELMGEGAASSNATLYRVVGDAALTQFDMCRVDTTAQVQKCLGKQLFVYIDPAYTNNTASGTGV
 30 GAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAI AVLHPTTIERNVNAAVEGNSSQDSGVA
 IATVLNEICPLPIHFLHYTDKSSALQWPIYMLGGEKSSAFETFIYALNSGTLSASQTVVSNNTIKISFDPV
 TYLVEQVRAIKCVPLRDGGQSYSAKQKMSDDLVAVVMMAHFMAATDDRMYKPISPQ
 >gi_4494933
 MLQKDAKLIFISSNSSDKSTSFLNLNKDAHEKMLNVVNYVCPDHKDDFNLQD TVVACPCYRLHIPAYIT
 35 IDETVRSTTNLFLEGAFTSTELMGDAATSAQSMHKIVSDSSLSQLDL CRVKSTSQDIQGAMKPC LHVYIDP
 AYTNTDASGTGIGAVIAVNHKVIKICILLGVEHFFLRDLTGTAAYQIASCAAALIRAI VTLHPQITHVNV
 AVEGNSSQDAGVAIATVLNEICSVPLSFLHHVDKNTLIRSPIYMLGPEKAKAFESFIYALNSGTFSASQT
 VVSHTIKLSFDPVAYLIDQIKAIRCIPLKDGHTYCAKQKMSDDVLVAAVMAHYMATNDK FVFKSLE
 >gi_7330018
 40 MLQKDAKLIFISSNSSDKSTSFLNLNKDAHEKMLNVVNYVCPDHKDDFNLQD TVVACPCYRLHIPAYIT
 IDETVRSTTNLFLEGAFTSTELMGDAATSAQSMHKIVSDSSLSQLDL CRVESTSQDIQGAMKPC LHVYIDP
 AYTNTDASGTGIGAVIAVNHKVIKICILLGVEHFFLRDLTGTAAYQIASCAAALIRAI VTLHPQITHVNV
 AVEGNSSQDAGVAIATVLNEICSVPLSFLHHADKNTLIRSPIYMLGPEKAKAFESFIYALNSGTFSASQT
 VVSHTIKLSFDPVAYLIDQIKAIRCIPLKDGHTYCAKQKMSDDVLVAAVMAHYMATNDK FVFKSLE
 45 >gi_4019255
 MLLLKAKALMENL TEASSTQSETEWTVDTPTMITNIKKSERMAYSKIGVIPSINLYSASLTSFCRLYRP

-17-

LALKQPLPQTGTLRLLPSEKPYISQKLSNYVKSLLTLKHVWHDIEAEAEYYASVQTEKTFMECPYIYLELRQ
 FIINLSSFLNGCYVKKSTHIEPFQLQLILHTFYFLISIKSPESTNKLFDIFKEYFGLGEMDSAMLQNFQKQ
 KASIFLIPIRRHGKTWIVVAIIISMLITSVENLHVGYVAHQKHVANSVFTEIINTLQKWFPSPKNIDVKKENG
 TIIYKIPGKKPSTLMCASC FNKNSIRGQTFNLLYIDEANFIKKDSLPAILGFMLQKDAKLIFISSVNSGD
 5 KATSFLENLKNASEKMLNIVNYICPDHKDDFSLQDSLISCPCYKLYIPTYITIDETIKNTTNLFLDGAFT
 TELMGDISVMSKNNIHKVIGETALMQFDLCRIDTTKPEITQCLNSIMYLYIDPAYTNNSEASGTGIGAI
 ALKNNSSKCIIVGIEHYFLKDLTGTATYQIASCACSLIRAALVLYPHIQAVHVAVEGNSSQDSAVAISTF
 LNECSPVKVNFMHYKDKTTAMQWPIYMLGSEKSAFESFIYAINSGTISASQSIISNTIKLTFDPISYLI
 EQIRAIRCYPLRDGSHTYCAKKRTVSDDVLVAVVMAHFFSTSNKHIFKQLNSI
 10 >gi_4019257
 MLQKDAKLIFISSVNSGDKATSFLENLKNASEKMLNIVNYICPDHKDDFSLQDSLISCPCYKLYIPTYIT
 IDETIKNTTNLFLDGAFTTELMGDISVMSKNNIHKVIGETALMQFDLCRIDTTKPEITQCLNSIMYLYID
 PAYTNNSEASGTGIGAIILKNNSSKCIIVGIEHYFLKDLTGTATYQIASCACSLIRAALVLYPHIQAVH
 VAVEGNSSQDSAVAISTFLENCSPVKVNFMHYKDKTTAMQWPIYMLGSEKSAFESFIYAINSGTISASQ
 15 SIISNTIKLTFDPISYLIEQIRAIRCYPLRDGSHTYCAKKRTVSDDVLVAVVMAHFFSTSNKHIFKQLNS
 I
 >gi_60355
 MLLLKAKKAIENLSEVSSTQAETDWMSTPTIITNTSKSERTAYSIGVIPSUNLYSSTLTSFCKLYHP
 LTLNQTQPTGTLRLLPHEKPLILQDLSNYVKLLTSQNVCHDTEANTEYNAAVQTQKTSMECPTYLELRQ
 20 FVINLSSFLNGCYVKRSTHIEPFQLQLILHTFYFLISIKSPESTNRLFDIFKEYFGLREMDPDMQIFKQ
 KASIFLIPIRRHGKTWIVVAIIISMLLTSVENIHVGYVAHQKHVANSVFTEIINTLQKWFPSPRYIDIKKENG
 TIIYKSPDKKPSTLMCATCFNKNSIRGQTFNLLYIDEANFIKKDSLPAILGFMLQKDAKLIFISSVNSGD
 RATSFLLENLKNASEKMLNIVNYICPDHKDDFSLQDSLISCPCYKLYIPTYITIDETIKNTTNLFLDGAFT
 TELMGDMGSGISKNNMHKVISMAITQFDLCRADTTKPEITQCLNSTMYIYIDPAYTNNSEASGTGIGAIL
 25 TFKNNSSKCIIVGMEHYFLKDLTGTATYQIASCACSLIRASLVLYPHIQCVHVAVEGNSSQDSAVAISTL
 INECSPKIVFYIHYKDKTTMQWPIYMLGAESIAFESFIYAINSGTISASQSIISNTIKLSFDPISYLI
 EQIRSIRCYPLRDGSHTYCAKKRTVSDDVLVAVVMAYFFATSNKHIFKPLNST
 >gi_695201
 MLQKDAKIIIFISSVNSSDQTTSFLYNLKNASEKMLNVVNYVCPQHREDFSLQESVVS CPCYRLHIPTYIA
 30 IDENIKDTTNLFMEGAFTTELMGDGAAATTQTNMHKVVGEPALVQFDLCRVDTGSPEAQRGLNPTLFLYV
 DPAYTNNTEASGTGMGAVVSMKNSDRCVVGVGEHFFLKELTGASSLQIASCAAALIRSLATLHPFVREAH
 VAIEGNSSQDSAVAIAITLLHERSPLPVKFLHHDKATGVQWPMYILGAEKARAFETFIYALNSNTLS CGQ
 AIVSNTIKLSFDPVAYLIEQIRAIRCYPLKDGTVSYCAKHKGSDDTLVAVVMAHYFATSDRHVFKNHMK
 QI
 35 >gi_4928934
 MLLSSFRNLQKNYKYSVQAQONIDWPVETPVLISKDSKTNRLAHPLIGVISRINLYSPTLKYYCDEYST
 TKQPKFTPDIGYVRDLKKHDQYFLPKLQHHLSTLCEAYNHVDRQAQVEFNASILTAKAFNANGVLNELKQ
 FLINLSCFLNGCYVSKSTCIELFQQLILHTFYFLISIKTPEETNKMFTFFKHVYGLFDIDDNMLQCFKQ
 KSTVFLIPRRHGKTWIVVAIIISVLLASVENVHIGYVAHQKHVANAVFTEIITTLYQWFPSPKNIEIKKENG
 40 TIIYTKPGRKPSTLMCATCFNKNSIRGQTFNILYVDEANFIKKEALPAILGFMLQKDAKIIIFISSVNSAD
 KSTSFLENLKNASEKMLNVVNYVCPHKKEDFNLQSTLTS CPCYRLHIPTYITIDESIKNTTNLFLDDVFT
 TELMGDISTFPTSSMFKVVEEQALFHFDICRVDTTQIDTVKIIDNVLYVYVDPAYTSNSEASGTGIGAVV
 PLKTKVKTIILGIEHYFLKNLTGTASQQIAYCVTSMIKAILTLHPHINHVNVAVEGNSSQDSAVAISTFI
 NEYCPVPVFFAHCNERSVFWQWPIYILGSEKSAFEKFCALNTGTLSASQTIVSNTIKISFDPVAYLME
 45 QIRAIRCLPLKDGSYTYCAKQKMSDDTLVAVVMANYMAISEKHTFKELCKT
 >gi_1632798

-18-

MLYASQGRGLTENLRNALQQDSTTQGCGLGAETPSIMYTGAKSDRWAHPLVGTI HASNLYCPMLRAYCRHY
 GPRPVFVASDESLEPMFGASPALHTPVQVQMCLLPCLRLDRLQRLPPNLEDSEALTEFKTSVSSARAILE
 DPNFLEMREFVTSLASFLSGQYKHKPARLEAFQKQVVLHSFYFLISIKSLEITDTMFDIFQSAFGLEEMT
 LEKLFHIFKQKASVFLIPRRHGKTWIVVAIISLILSNLSNVQIGYVAHQKHVASAVFTEIIDTLTKSFDK
 5 RVEVNKETSTITFRHSGKISSTVMCATCFNKNSIRGQTFHLLFVDEANFIKKEALPAILGFMLQDKAKII
 FISSVNSADQATSFLYKCLKDAQERLLNVVSVCQEHQDFDMQDSMVSCPCFRLHIPSYITMDSNIRATT
 NLFLDGAFSTELMGDTSSLSQGSLSRTVRDDAINQLELCRVDTLNPRVAGRLASSLYVYVDPAYTNNTSA
 SGTGIAAVTHDRADPNRVIIVLGLHEHFFLKDLTGDAALQIATCVVALVSSIVTLHPHLEEVKVAVEGNSSQ
 DSAVAIASIIGESCPLPCAFVHTKDKTSSLQWPMYLLTNEKSKAERLIYAVNTASLSASQVTVSNTIQL
 10 SFDPVLYLISQIRAIKPIPLRDGTYTYTGKQKNLSDDVLVALVMAHFLATTQKHTFKKVH
 >gi_2337991
 MFYVKVMPALQKACEELQNWSAKSGKWPVPETPLVAVETRRESERWPHPYLGLLPGVAAYSSTLEDYCHL
 YNPYIDALTRCDLGQTHRRVATQPVLSQDLQQLKFLSCPRNTSVKAKLEFEAAVRTHQALDNSQVFLE
 LKTFVLNLNSAFLNKRYSRSSHIEFLQKQLIMHTFFFLVSIKAPELCEKFCNIFKLYFNIDTMDQATLDI
 15 FKQKASVFLIPRRHGKTWIVVAIISILLASVQDLRIGYVAHQKHVANAVFTEVINLTHTFFPGKYMDVKK
 ENGTIIFGLPNKKPSTLLCATCFNKNSIRGQTFQLLFVDEANFIKDALEPTILGFMLQDKAKIIFISSN
 SSDQSTSFLYNLKGASERMLNVVSVCNHNKEDFSMQDGLISPCYSLHVPYSYISIDEQIKTTTNLFLDG
 VFDTELMGDSSCGTLSTFQIISSEALSQFELCRIDTASPOVQAHNLSTVHMYIDPAFTNNLDASGTGISV
 IGRLGAKTKVILGCEHFFLQKLTGTAAALQIASCATSLLRSVVIHPIKCAQITIEGNSQDSAVAIANF
 20 IDECAPIPVTFYHQSDKTKGVLCPLYLLGQEKAVAFESFIYAMNLGLCKASQLIVSHTIKLSFDPVTYLL
 EQVRAIKQSLRDGSHYHAKQKNLSDDLVSVMMSLYLSSANTLPFKPLHIERFF
 >gi_2317977
 MLQKDAKIFFISSVNSGEKTTSTFLYNLKDANEKMNVSVCSEHMEFNNKQSAITACPCYRLVPEFIT
 INDNIKCTTNLLLEGSFATELMGNMQSHTEVSGNSMIHSSSLTRLDYRCDTAGQGAPTENTLFFVYIDP
 25 AYGNNVHASGTGIVAMSHCKHTKKCIILGLEHFFLNLTGTAAHNIASCATALLEGILFQHPWIQEIRCI
 IEGNSNQDSAVAIATFISHNIKLPFLFASYRDKTGMQWPIYMLSGDKTLAFQNFISSLNQGLLCASQTVV
 SNTVLLSSDPISYLIEQIKNTKCIYHKNKTITTFQSKTHTMSDDVLIACVMTCYVMTTNKISYISFSIK
 >gi_6625593
 MFIAKKSYFEAVRSTVSSHSEEFWKSDDPVYFTQYKKQCNRLPNAYLGTLSASKYSENFRHYVATFS
 30 NSPLDFPQSVFNERNPCEYSVPYLDALQCSAKTLVGCSVSTTERNEYEVCKEATRCFKDAMSHKVLKVF
 LSNLSWFLKGHYKSKQAFLEPFQKQLILHSFMFVASIKCPETTTKLFDEFKFLDMLYFDNTDLLTFLQK
 SPAFLIPRRHGKTWIVTAIISMLLTSVDDLHIGYVAHQKHVSLAVFLEISNILLAWFPRKNIDIKKENG
 ILYSHPGKKSSTLMCATCFNKNSIRGQTFNLLFVDEANFIKKEALPAILGFMLQDKAKIFFISSVNSGEK
 TTSFLYNLKDANEKMNVSVCSEHMEFNNKQSAITACPCYRLVPEFITINDNIKCTTNLLLEGSFAT
 35 ELMGNMQSHTEVSGNSMIHSSSLTRLDYRCDTAGQGAPTENTLFFVYIDPAYGNNVHASGTGIVAMSHC
 KHTKKCIILGLEHFFLNLTGTAAHNIASCATALLEGILFQHPWIQEIRCIIEGNSNQDSAVAIATFISH
 NIKLPFLFASYRDKTGMQWPIYMLSGDKTLAFQNFISSLNQGLLCASQTVVSNVLLSSDPISYLIEQIK
 NTKCIYHKNKTITTFQSKTHTMSDDVLIACVMTCYVMTTNKISYISFSIK

gi_10180719	MFGLLGEETKRHFERLAKTKNDRLGASHNRNERSIRGG---DHVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP	1	130
gi_17673189	MFGLLGEETKRHFERLAKTKNDRLGASHNRNERSIRGG---DHVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_5689285	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_1869837	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_59501	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_2605992	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_330792	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_971317	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_5869808	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_5708110	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_1813970	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_2748296	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_325496	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_5733564	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_1136808	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_2246552	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_4019255	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_50355	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_4928934	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_2337991	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_1632798	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_6625593	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_1718281	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_2246515	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_4494933	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_7330018	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_4019257	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_695201	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_2317977	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_854039	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
gi_4996048	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		
Consensus	MFGLLGEESAKKHAFERLLRDRNERLGASRKNELARGG---SLVOAPF---LNFATPVRPRHQTVPATGILHNCCDSLGTYSAITTRALYSSIRACSEFDELRRD---SYPRCYP		

TABLE 2

gi_10180719	ERLLRVAFDTPHFSDIVTRHFQRATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	261
gi_7673189	ERLLRVAFDTPHFSDIVTRHFQRATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	270
gi_5689285	ERLLRVAFDTPHFSDIVTRHFQRATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	280
gi_1869837	NTFLRLVFEIPLFSDIVTRHFQRATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	290
gi_59501	NTFLRLVFEIPLFSDIVTRHFQRATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	300
gi_2605992	GAFLKRVFNTPEFSOATIRHFQRATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	310
gi_330792	GAFLKRVFNTPEFSOATIRHFQRATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	320
gi_5869808	GAFLKRVFNTPEFSOATIRHFQRATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	330
gi_5708110	DCFLRTVFNTPSVDSVLEHFKQKSTVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	340
gi_1813970	DCFLRTVFNTPSVDSVLEHFKQKSTVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	350
gi_2746296	NRVLSVFNISLSENIQHFQRITVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	360
gi_325496	LOVLTAFQIDFELSDIIFKQKATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	370
gi_1136808	LOVLTAFQIDFELSDIIFKQKATVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	380
gi_2246552	INVLTHVDFEFVNESTLNLKQKTNVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	390
gi_4013255	INVLTHVDFEFVNESTLNLKQKTNVFLVPRRHGKTHFLVPLIALAHSSFEIGIRIGYTSHTRKIEPVFEIDIGRLRHF GAHRYDVHVKG-ITITFSFSGLSITVTFASSHNTN-SIRGQDFNLLFVD	
gi_4928934	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_2337991	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_1632798	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_6625593	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_1718281	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_2246515	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_4494933	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_7330018	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_4013257	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_695201	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_2317977	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_854039	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
gi_4996040	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	
Consensus	FGLFKQYFGLFETPNVSVLQTFKQKASVFLIPRRHGKTHVVAITISMLLSVENINIGYVVAHQKHAVSVFTEIINTLQKAPSKNIDVYKKEGIIYKIPORPSILLCHSCFHNK-SIRGQTFNLLYD	

TABLE 2 CONTINUED

gi_10180719	391	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	520
gi_7673189	400	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	510
gi_5689285	410	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	500
gi_1869837	420	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	490
gi_59501	430	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	480
gi_2605992	440	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	470
gi_330792	450	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	460
gi_971317	460	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	450
gi_5869808	470	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	440
gi_5708110	480	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	430
gi_1813970	490	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	420
gi_2746296	500	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	410
gi_325496	510	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	400
gi_5733564	520	EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	391
gi_1136808		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_2246552		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_4019255		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_60355		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_4928934		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_2337991		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_1632798		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_6625593		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_1718281		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_2246515		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_4494933		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_7330018		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_4019257		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_635201		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_2317977		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_854039		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
gi_4996048		EANFIRPDVAQVITIGFLNQATCKITFVSSNWSGKASTSFLYGLKGSODDLNVTYTCDEHKKHNTDTHATSCSYLHKPVFIITHDGHARRTAELFLPDSFQHEIIGGYYVORTICQGRSIFIASAI	
Consensus	nlqkdak, ifiss.ns.d.stcsfl,nlkna, ek\$lnvv\$yic.#h..df..q....aCpCyrLh.P.%Iki#..jk..It.Hlf\$egaf..lElmg.. ..a.....t.e.al	

TABLE 2 CONTINUED

gi_10180719
gi_17673189
gi_15689285
gi_1863937
gi_153501
gi_12605992
gi_1330792
gi_1971371
gi_15863908
gi_15708110
gi_1813970
gi_12748296
gi_1325496
gi_15733564
gi_1136800
gi_1246552
gi_14013255
gi_160355
gi_4928934
gi_12337991
gi_11632798
gi_1625593
gi_1178281
gi_12465515
gi_4494933
gi_17330018
gi_14019257
gi_695201
gi_12317977
gi_854039
gi_4956048
Consensus

TABLE 2 CONTINUED

gi_10180719	781	791
gi_7673189	-----+	
gi_5689285	VEATCRGPTER	
gi_1869837	TS	
gi_59501	VS	
gi_2605992	q	
gi_330792	q	
gi_971317	RPPAD	
gi_5869808		
gi_5708110	IGRRQRP	
gi_1813970		
gi_2746296		
gi_325496		
gi_5733564		
gi_1136808		
gi_2246552		
gi_4019255		
gi_60355		
gi_4928934		
gi_2337991	FF	
gi_1632798		
gi_6625593		
gi_1718281		
gi_2246515		
gi_4494933		
gi_7330018		
gi_4019257		
gi_695201		
gi_2317977		
gi_854039		
gi_4986048		
Consensus		**

TABLE 2 CONTINUED

-26-

Table 3. Degenerate primers generated by CODEHOP

Block x7263xbliD

T L Y V Y I D P
 oligo:5'-AACCTGTACGTGtayntngaycc-3' degen=64 temp=33.4 Extend clamp

T L Y V Y I D P A
 oligo:5'-AACCTGTACGTGTACntngayccngc-3' degen=128 temp=36.0 Extend clamp

T L Y V Y I D P A Y
 oligo:5'-AACCTGTACGTGTACATngayccngcnt-3' degen=128 temp=42.5 Extend clamp

Complement of Block x7263xbliD

Y I D P A Y T N N T
 atrnanctrggGCGGATGTGGTTGTTGT oligo:5'-TGTTGTTGGTGTAGGCGggrtcnanrta-3'
 degen=64 temp=62.9

D P A Y T N N T R A
 anctrngcngnaTGTGGTTGTTGTGGGTCCG oligo:5'-GCCTGGGTGTTGTTGGTGTangcnggrtcna-3'
 degen=128 temp=61.8

D P A Y T N N T R A
 ctrngcngnawGTGGTTGTTGTGGGTCCG oligo:5'-GCCTGGGTGTTGTTGGTGWangcnggrtc-3'
 degen=64 temp=61.0

Block x7263xbliE

C I I F G M E H F F
 oligo:5'-TGGATCATCTTCGGCATngarcaytwyt-3' degen=64 temp=55.7 Extend clamp

I F G M E H F F L
 oligo:5'-CATCTTCGGCATGGAGcaytwytwyyt-3' degen=64 temp=62.0

Complement of Block x7263xbliE

E H F F L R D L T G
 ctygtrawrawGGACTTCCTGGACTGCCC oligo:5'-CCCGTCAGGTCCTTCAGGwarwartgytc-3'
 degen=32 temp=61.7

H F F L R D L T G
 tygtrawrawrrrACTTCCTGGACTGCCCC oligo:5'-GCCCCTCAGGTCCTTCarrwarwartgyt-3'
 degen=128 temp=60.8

H F F L R D L T G
 gtrawrawrrraCTTCCTGGACTGCCCC oligo:5'-GCCCCTCAGGTCCTTCarrwarwartg-3' degen=64
 temp=60.8

Block x7263xbliF

E V H I A V E G N
 oligo:5'-GGACGTGCACGTCGCCrtnrgarggnaa-3' degen=64 temp=63.8

Complement of Block x7263xbliF

E G N S S Q D S A
 anctyccnttrwGGTTGGTCCTGAGGCGG oligo:5'-GGCGGAGTCCTGGTTGGwrttnccytcna-3'
 degen=128 temp=62.7

E G N S S Q D S A V
 ctyccnttrwsGTTGGTCCTGAGGCGGC oligo:5'-CGGCGGAGTCCTGGTTGswrttnccytc-3'
 degen=64 temp=63.9

-27-

CLAIMS

1. A method of designing a panel of degenerate primer pairs for screening for new members of multiple known virus families in a biological sample, wherein each primer pair in the panel binds a sequence that is conserved across members of a said virus family and selectively directs amplification of sequence of said family by PCR,
5 which method comprises

(a) providing a plurality of amino acid sequences from members of a first virus family,

10

(b) comparing the sequences to identify conserved regions,

(c) designing a first primer pair using a computer based method, wherein each primer in the pair binds a nucleotide sequence that encodes a conserved region identified in

15

(b) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the nucleotide sequences that encode conserved regions in members of the first virus family, and

(d) repeating steps (a) to (c) for each virus family.

20

2. A method of designing a panel of degenerate primer pairs for screening for new members of multiple known virus families in a biological sample, wherein each primer pair in the panel binds a sequence that is conserved across members of a said virus family and selectively directs amplification of sequence of said family by PCR,
25 which method comprises

(a) providing a plurality of nucleotide sequences from members of a first virus family,

30

(b) comparing the sequences to identify conserved regions,

-28-

(c) designing a first primer pair using a computer based method, wherein each primer in the pair binds a conserved region identified in (b) and wherein the primer pair is designed to amplify by PCR the nucleotide sequence between the conserved regions in members of the first virus family, and

5

(d) repeating steps (a) to (c) for each virus family.

3. A method according to claim 1 or 2 which further comprises synthesising one or more of the primer pairs and determining optimal conditions for using the primer pairs in PCR.

10

4. A method according to any one of the preceding claims which comprises testing the ability of one or more of the primer pairs to amplify a nucleotide sequence that encodes an amino acid as defined in claim 1(a) or a nucleotide sequence as defined in claim 2(a).

15

5. A method according to claim 3 or 4 which comprises testing the primer pair(s) in a range of buffer conditions to determine the optimal buffer conditions for PCR.

20 6. A method according to any one of claims 3 to 5 which comprises testing the primer pair(s) at a range of different temperatures to determine the optimal temperature for PCR.

7. A method according to any one of the preceding claims which comprises identifying one or more groups of primer pairs wherein the primer pairs in each group have similar optimal conditions of use in PCR such that they can be used optimally in the same reaction vessel.

25

8. A method according to claim 7 wherein each primer pair in a group generates a PCR product of a different size to the other primer pair(s) in the group.

30

-29-

9. A method according to claim 7 or 8 wherein each primer pair in a group carries a different label from the other primer pair(s) in the group.
10. A method according to claim 9 wherein each primer pair in a group carries a
5 differently-coloured fluorescent label.
11. A method according to any one of the preceding claims wherein the biological sample is a single-source sample from a single individual or is a pooled sample from more than one individual of the same species.
- 10 12. A method according to claim 11 wherein the biological sample is a human sample.
13. A method according to any one of the preceding claims wherein at least 50% of
15 the primer pairs bind a sequence that is conserved across all of the genres and/or subfamilies.
14. A panel of primers designed according to any one of the preceding claims.

FIG 1

1/1

